DESIGN AND IMPLEMENTATION OF COVID-19 PREDICTION BY USING MACHINE LEARNING MODEL

(The synopsis submitted towards the partial fulfilment of requirement for the award of degree of)

P.hD

In

ELECTRONICS AND COMMUNICATION ENGINEERING

Under the guidance of
Dr. Gurinder Kaur Sodhi
HOD & Assistant Professor
Electronics & Communication Engineering

Submitted by
Deepti Malhotra
Roll No. 19380179001
Research Scholar

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING
Desh Bhagat University, Mandi Gobindgarh
Punjab, India

Candidate's Declaration

I hereby declare that the thesis entitled "**Design and Implementation of Covid-19 prediction by using Machine Learning Model**" submitted by me in partial fulfilment of requirements for the award of degree of Ph.D (Electronics and Communication) to Desh Bhagat University is a record of my own work carried out under the supervision and guidance of **Dr.Gurinder Kaur Sodhi, HOD & Assistant Professor, DBU.**

To the best of my knowledge this thesis has not been submitted to earlier to any other University or Institute for the award of any degree.

Date: **Deepti Malhotra**
Place: DBU, Mandi Gobindgarh **Roll no. 19380179001**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of supervisor

Dr Gurinder Kaur Sodhi
HOD & A.P, DBU

CONTENTS

	Title	**(i)**
	Declaration	**(ii)**
	Table of Contents	**(iii)**
Chapter 1	**Introduction**	**1-7**
Chapter 2	**Extracts of Literature Search**	**8-13**
2.2	Literature Review	8
2.3	Research Gaps	13
Chapter 3	**Problem Formulation and Objectives**	**14**
3.1	Problem Formulation	14
3.2	Objectives	14
Chapter 4	**Methodology**	**15-16**
Chapter 5	**Scope of Study and Possible Developments**	**17**
5.1	Scope of study	17
5.2	Expected Developments	17
	References	

Chapter 1

Introduction

1.1 Corona virus disease 2019 (COVID-19)

The coronavirus originated in Chinese province, Wuhan in December 2019. As of January 2021, more than 95 million cases have been reported worldwide, with a death rate of 2% of the overall closed cases. This fast spread of the pandemic characterizes a universal challenge and poses a grave danger to people health and the economy around the world. The coronavirus disease 2019 is a virus of the coronavirus family. Studies suggest that COVID-19 encompasses medicaltra its similar to SARS-CoV [1]. Major symptoms are fever and cough, while gastrointestinal signs are rare. The absence of fever in novel coronavirus- infected people is more common than in patients diseased with the same virus, that is, MERS corona virus (2%) and SARS corona virus (1%). Thus, non-febrile patients are likely to be exempted from a monitoring device with the main attention on diagnosing fever. According to reports, early patients diseased with COVID-19 showed a relation to a big seafood and animal market in Wuhan, which established an animal-to-person transmission. In contrast, a large number of patients showed no association with animal markets, suggesting the actuality of human-to-human transmission of coronavirus disease 2019.

1.1.1 Machine learning

The healthcare sector is a huge sector requiring medical data whose collection and processing is required in real-time. Furthermore, the core of this sector faces the challenge of data management which needs ground-level prediction and distribution of information for physicians to seek prompt medical care. Dominant aspects in this sector, including practitioners and health-based organizations etc., have made an effort to gather, handle, and regenerate data to facilitate improvement in medical practices and using it for technical revolution. Nevertheless, several factors such as presence of huge amount of data, security concerns etc. have made it quite difficult to deal with medical data.

The emergence of AI has proved to be a technological breakthrough of the 21th century. It includes many applications in areas ranging from weather forecasting, astrophysical investigation to self-directed systems. Machine learning based prediction has gained huge

admiration in the last few years. The course of learning is basically learning from past work experience or observations, such as instances, or instructions, to see patterns in data. This process makes use of examples to improve the decision-making ability of the system. The fundamental objective of machine learning is for the computer to automatically learn and adjust the performance activities without any human intervention [2].

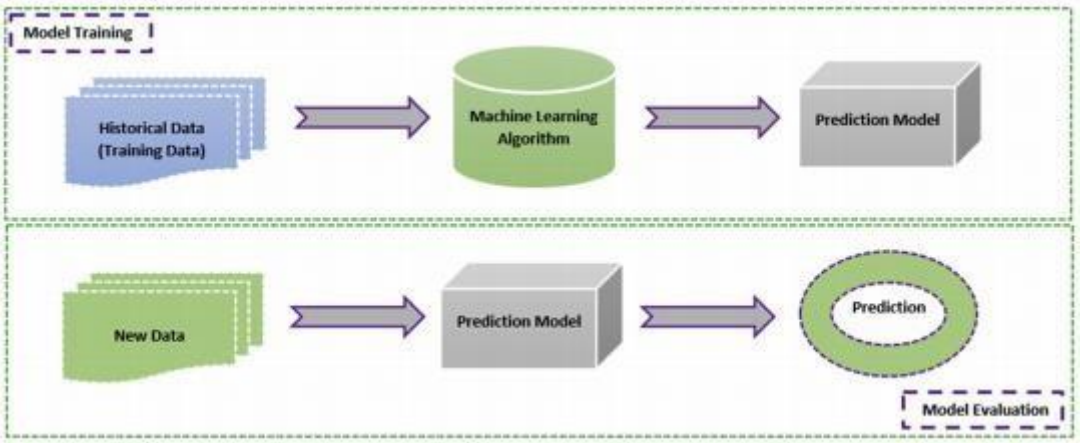

Fig.1.1: Machine learning Process

Fig.1.1 depicts the methodology of machine learning. Using this previous data, the training of the model is performed. Next, the new data is tested using trained system and then make predictions. The performance of the trained ML technique is evaluated using some part of the previous data existing (which is absent during training). This process is commonly known as the verification. This process involves evaluating the machine learning framework for its performance measure, such as accuracy. Accuracy defines the performance of ML models on hidden data with regard to various attributes correctly predicted and the proportion of the overall features existent to be predicted.

1.2 Covid-19 Disease Prediction using Machine Learning

The epidemiological data containing the number of people who suffered with this disease previously and the total population is employed in the predictive models of corona virus 2019. Different parameters namely latency period and probability of healing are utilized for predicting the infection trends. But, these models are inefficient to reflect diverse socio-static and economic factors due to which the course of the virus is affected. Thus, it is important to predict the disease on the basis of economic and social data and to analyze the trends of COVID-19 on the basis of epidemiological data.

Fig. 1.2 represents the generalized COVID-19 prediction model based on data mining.

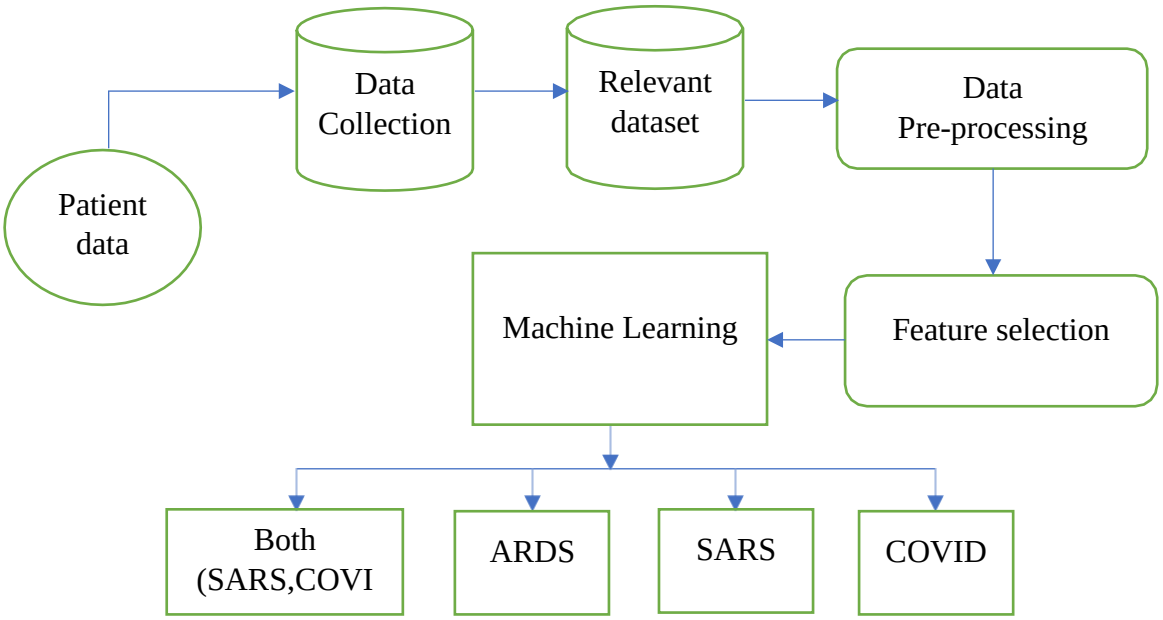

Fig. 1.2: General Composition of COVID-19 prediction model

All stages of this predictive framework have been explained below:

a. **Data collection**: The Novel Coronavirus 2019 is confirmed as a medical emergency by WHO. An open access is provided by the researchers and hospitals to the data of this contagion [3]. Various sources are utilized to gather the data. The GitHub, open-source data repository is one among them. The data of 212 patients detected with symptoms of corona virus and other viruses is comprised in this dataset.

b. **Relevant dataset**: There are certain confounding factors included in the gathered data because of which prediction of corona disease becomes difficult. Generally, the datasets of corona disease are highly imbalanced that lay negative impact on predictive accuracy. The sampling techniques, cost-sensitive learning and feature selection are extensively implemented by the statisticians for carrying the predictions with the help of highly imbalanced datasets. But, there is not any guideline provided for utilizing the technique and predictive system based on the nature of the dataset. The labels are assigned to the clinical datasets to their corresponding classes. Most of the datasets are classified in 4 classes such as COVID, ARDS, SARS and Both (COVID, ARDS).

c. **Pre-processing:** In this process, the training data is sampled to be suitable for classification algorithm. The pre-processing method acts effectively in order to tackle an enormous volume of data or in cases at which domain knowledge is deficient. There are 3 stages included in the process of preparing the data in which data is handled, outliers are removed and data is balanced. The missing values are handled using the KNN Imputer in which the KNN (k-nearest neighbours) algorithm is deployed for imputing the missing values. To achieve this, the mean value is utilized from nearest neighbours. The anomalies present in a given dataset are detected by the means of Outlier detection technique. There is a difference between the anomalies and the normal records with regard to the quantity and quality. Thus, the performance of classifier is enhanced with the elimination of outliers [4].

d. **Feature Selection**: This is an essential procedure to construct a ML framework. The selection of suitable attributes is useful to alleviate the data redundancy and to avoid the noisy data. Consequently, the performance of model is enhanced. The attributes based on filter are selected using Relief feature selection. In every iteration, an instance is chosen from the available instances using this algorithm in the dataset at random. Subsequently, feature relevance is updated on the basis of difference amid selected instances and two close neighbour instances.

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \quad \ldots.(1)$$

This system assists in selecting 18 attributes on the basis of their significance to detect the corona virus 2019 on the basis of clinical studies and reported in previous ML prediction models. Haemoglobin, Platelets, Leukocytes, Neutrophils, Age, Urea are some of these attributes.

e. **Machine learning**: The ML (machine learning) is emerged as a potential technology as the uncertainties and a high degree of complexity are occurred in the advancement of epidemiological models of the outbreak. The efficient outcomes are obtained from the machine learning for constructing the SIR (Susceptible-Infected-Recovered) model that performed more efficiently and proved robust and capable of generalization. ML is considered as the computing method having great potential for predicting the corona virus disease.

Following are the some commonly used Covid-19 outbreak prediction models.

i. **K-Nearest Neighbours Algorithm**: It is a non-parametric approach. Thomas Cover, an engineer, coined it to perform tasks of classification and regression. During classification, the output represents a class membership. To classify an object, this algorithm depends on the plurality vote of its nearby instances. The allocation of an object is done to the class which most common among its nearest neighbors (k is a positive integer, usually small). If k = 1, the object is allocated only to the class of that single nearest neighbor. In regression, the output denotes the object's feature value. This value represents the average of the values of k nearest neighbors [5].

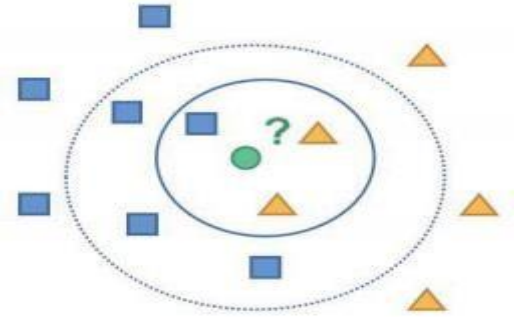

Fig.1.3: Example of k-NN classification

As shown in Fig.1.3, it is required to classify the test instance (green dot) into either blue squares or yellow triangles. When k = 3 (solid line circle) it is allocated to the yellow triangles as there exist 2 triangles and just 1 square within the innermost circle. When k = 6 (dashed line circle) it is allocated to the blue squares (4 squares inside the outer circle vs. 2 triangles).

ii. **Logistic Growth Model:** It is firstly almost exponential, but has a lower rate of growth as the output is inclined towards the upper bound of the model, known as the carrying ability. The model represents the logistic growth of the population with time for three constants called a, b, and c:

$$f(x) = \frac{c}{1 + ae^{-bx}} \quad \dots\dots(2)$$

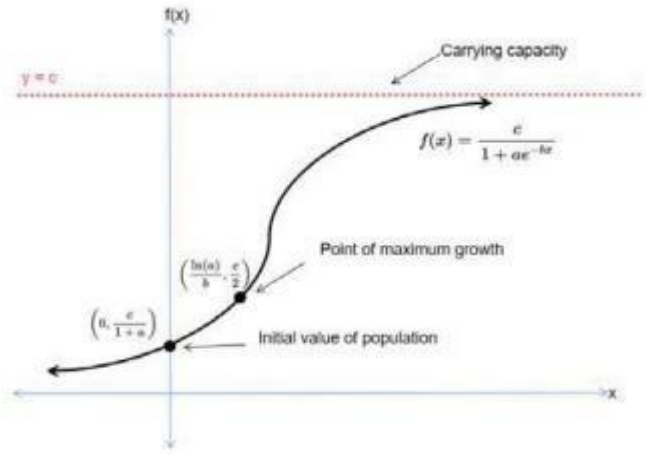

Fig. 1.4: The growth rate of logistic growth model

Fig.1.4 reflects the change in growth rate over time. The graph moves upward from left to right. However, the growth rate only moves upward until it reaches its point of maximum growth rate, at which point the rate of growth goes down [6].

iii. **Long Short-Term Memory:** It is an efficient algorithm for predicting the time series data. The most significant attribute of this algorithm is the integration of earlier output with current input for utilizing the neurons. A time window was included in this algorithm. The major goal of this algorithm was to vanish the gradient expose it in long time series training. The figure 2 represents a typical Long Short-Term Memory neuron.

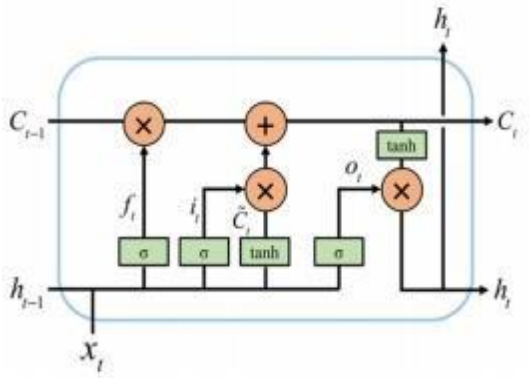

Fig.1.5: A typical LSTM neuron

LSTM has potential to learn the long-term dependencies. This attribute based on a special structure known as Cell state or a rather sophisticated gate control system. Given the first equation is input gate, second is forget gate and the third equation is the output gate.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \ldots\ldots (3)$$

$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big) \ldots. (4)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \ldots. (5)$$

The forget gate is a layer which contains a sigmoid for controlling whether the earlier information is utilized in the current neuron [7]. The current information is decided by the input gate. Sigmoid is useful in selecting the suitable information to be updated and tanh is employed for generating a novel candidate memory whose selection is done with the sigmoid. The output gate assists in providing the value for rest of the neurons. These gates are useful for LSTM in generating long memory and short memory with the given equation for a long or short memory.

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \ldots (6)$$

$$h_t = o_t \tanh C_t \ldots. (7)$$

The network attains potential for dealing with the long-term dependency when the value of finite data is increased in a long range.

Chapter 2

Literature Review

Pratima Kumari, et.al (2020) intended ML (machine learning) method called ANN in order to predict the Coronavirus Disease 2019 outbreak in India for the first time [8]. In addition, a mathematical curve fitting system was adopted for determining the performance of the intended method. The development of the epidemic at diverse transmission rates for analyzing the impact of preventive measures such as lockdown and social distancing. The intended method was proved appropriate to predict the development rate of metrics of corona disease at lower MAPE values and cumulative deceased cases.

Yasin Khan, et.al (2020) formulated a CNN algorithm to carry out a multi-class classification of the chest X-ray images of Coronavirus Disease 2019 as infected patients, patients of pneumonia and healthy persons [9]. Moreover, the Monte-Carlo simulation was utilized on the original data distributions for improving the data so that the disease was predicted. Moreover, COVID-19 pandemic was forecasted using LR of the components of GMM. The formulated algorithm was quantified on the Kaggle and the University of Montreal dataset that comprised chest X-ray pneumonia images. The accuracy obtained from the formulated algorithm was computed 100% in training phase and 96.66% in the testing phase.

Choujun Zhan, et.al (2021) projected a ML (machine learning) mechanism in order to forecast the COVID-19 on the basis of BLS (Broad Learning System) [10]. The significant attributes were extracted using RF (Random Forest). Therefore, RF-Bagging-BLS (Random-forest-Bagging Broad Learning System) was constructed by integrating the bagging strategy with BLS with the objective of predicting the COVID-19 pandemic. The projected mechanism performed more effectively with regard to RMSE (relative mean square error), MAD and MAPE in comparison with other models. Thus, the projected mechanism outperformed the traditional models.

Yifan Yang, et.al (2020) suggested a LSTM algorithm to predict the infected population in China [11]. But, this algorithm was incapable of defining the dynamics of diffusion procedure and provided higher long-term prediction error. Thus, an extensively utilized model named SEIR (susceptible-exposed-infected-removed) was put forward for capturing the spread process of Coronavirus Disease 2019. A sliding window technique was

implemented to estimate the parameter and predict the infected populations. The suggested algorithm offered some insights for epidemiological studies and to understand the transmission rate of the current COVID-19.

Leonardo Sestrem de Oliveira, et.al (2021) introduced an ANN with the objective of predicting the number of Coronavirus Disease 2019 concerning confirmed cases and deaths and also the future seven days for the time series of Brazil, Portugal and the United States [12]. The simulation results depicted that the introduced method was capable to predict the confirmed cases and deaths from Coronavirus Disease 2019. The ANN utilized with a specific test set provided 50% higher MSE as compared to the Artificial Neural Network with a random test set. The least MSE was acquired for all scenarios when the sigmoidal and linear activation functions were integrated with the LM function.

Durga Prasad Kavadi, et.al (2020) recommended a PDR-NML technique to predict Coronavirus Disease 2019 which was spread all over the world [13]. The optimal metrics were investigated efficiently using PPDLR model. Subsequently, the NML was adopted for the normalized attributes so as the accurate predictions were achieved. The results indicated the superiority of the recommended technique over existing schemes in the Indian population and it was presented as a promising tool to make predictions for other countries.

Ankan Ghosh Dastider, et.al (2021) investigated a CNN (convolutional neural network) algorithm in which the fusion of an autoencoder network and separable convolutional branches was done with a modified DenseNet-201 network [14]. Hence, a vigorous, noise-free classification system was constructed. The accuracy of investigated algorithm was evaluated using a five-fold cross-validation system. The in-depth result analysis proved that when LSTM layers were utilized, the classification performance was enhanced. This algorithm provided performed more effectively as compared to conventional techniques. The extensive analysis confirmed the applicability of the investigated algorithm to detect the severity scores of Coronavirus Disease 2019 from LUS images.

Abhishek Dixit, et.al (2021) established a 3-step approach that employed K-means clustering and extracted the attributes in the data pre-processing stage [15]. The subsequent stage focused on optimizing the selected attributes. For this, a novel feature optimization technique was deployed on the basis of hybrid DEA (differential evolution algorithm) and PSO. Thereafter, the SVM made the utilization of optimized attributes. The empirical outcomes

demonstrated that the established approach yielded the accuracy around 99.34% and their robustness and adaptability of this approach was also proved in detecting the patients suffered from COVID-19.

Elena Casiraghi, et.al (2020) constructed a computerized mechanism whose major intend was to extract the most relevant radiological and laboratory variables for improving the process of predicting the patient risk [16]. After that, an adequate ML (machine learning) model was put forward. A robust and consistent variable was selected by integrating Boruta with RF (Random Forest) in a 10-fold cross-validation system for achieving a variable importance estimate that was impartial when the surrogates were available. The selection of significant variables was done for training the RF algorithm whose robustness and efficacy was presented. The implementation of the developed system was easy and assisted the emergency departments in predicting the rapid and accurate risk in COVID-19 patients.

Anwar Jarndal, et.al (2020) discussed that a model was formulated for estimating the number of deaths in the affected cases on the basis of documented number of older diabetic and smoking cases [17]. The GPR algorithm was adopted for building a model. A comparative analysis was done on this algorithm against ANN (Artificial Neural Network) model. A trustworthy data gathered from WHO for distinct countries was utilized in this algorithm. The adopted algorithm was efficient to estimate the number of deaths because of any arbitrary number of inputs. The effective measures were planned for mitigating the number of deaths using this algorithm.

Huan Zhao, et.al (2021) designed the BP (back propagation neural network) algorithm for analyzing and predicting the growth rate of Coronavirus Disease 2019 [18]. Initially, the gathering of epidemic data of Italy was done from WHO. Subsequently, R and MATALB were employed to analyze a number of indicators. A computer language was executed to put together the results obtained from multiple indicators for predicting the COVID-19 in Italy. The predictive outcomes indicated that the trend of Italy epidemic situation in the study period was found good. But, the measures were required to prevent and control the corona disease. Eventually, a comparison of fitting value of the designed algorithm was done with real value and the fitting degree was obtained upto 0.99.

Nanning Zheng, et.al (2020) suggested a hybrid AI (artificial-intelligence) technique in order to forecast the Coronavirus Disease 2019 [19]. At first, an ISI (improved susceptible-

infected) model was presented for estimating several infection rates so that the transmission laws and the growth trend were analyzed. Later, the effects of measures taken to prevent and control the disease were taken in account for embedding the NLP (natural language processing) and LSTM network into the presented model with the objective of developing the suggested hybrid technique to predict coronavirus disease. The outcomes obtained in experimentation depicted that the suggested technique was capable of mitigating the errors of the predictive results.

Safa Bahri, et.al (2020) intended a LSTM (Long Short-Term Memory)-based predictive framework to predict the future cumulative number of recovered cases of CoronaVirus Disease 2019 seven days earlier, in India, USA and Italy [20]. This framework was suitable for the dynamic of evolution of recovered cases of the observed Coronavirus Disease 2019 and predicting the disease accurately with a total error within 1.46 to 2.65%. The experimental results revealed the effectiveness of intended framework which provided a minimal error below 3%. This model was proved as an effective tool for practitioners for monitoring the occurrence of corona disease all over the world.

Teddy Mantoro, et.al (2020) introduced two techniques namely SVR and SIR system for predicting COVID-19 [21]. The disease was predicted in the best-case as well as worst-case scenario. The initial scenario included the current daily case as a maximum case and another country's maximum case were contained in the latter scenario utilized. The SVR algorithm provided diverse end of epidemic. The SIR model was utilized to validate the similar end of epidemic on January 2021 in both cases. However, a rapid increase of infectious people out of 450,000 people was found in initial scenario to 5,500,000 people in latter scenario. The prediction obtained using the introduced approach offered an insight for the policy maker to deal with the coronavirus disease pandemic.

B. Prabha, et.al (2021) constructed a novel ML (machine learning) method known as HDDP (Hybrid Disease Detection Principle) which was developed using two traditional ML model namely CNN (Convolutional Neural Network) and the AdaBoost Classifier [22]. The ML factors were considered for the CT image of lung that was pre-processed, its attributes were extracted and the classification was done. The constructed method had potential for handling the CT images on the basis of these effective image processing strategies. Hence, the COVID-19 disease was predicted and a report was sent to the respective user with proper

accuracy ratio. This approach focused on predicting the disease effectively and assisted the doctors and patients successfully.

Wencheng Zou, et.al (2020) devised the ML (machine learning) techniques namely LR (linear regression) and NN (neuron networks) to predict the number of positive cases of Coronavirus Disease 2019 [23]. The LR algorithm was deployed for gathering the data of a state so that the corona disease was predicted. The number of infections was predicted using the data taken from Georgia and Massachusetts. The dataset was split into3 consecutive time periods and diverse models were trained to be suitable for each corresponding data for comparing MSE (mean square error) values. Moreover, the general viewpoint despite of three time periods depicted the superiority of single variable linear regression (LR) over accurately in comparison with the traditional technique with regard to accuracy.

Hamza Turabieh, et.al (2021) established a hybrid approach in which the wrapper feature selection named GA (genetic algorithm) and various classification algorithms for determining the occurrence of Coronavirus Disease 2019 on the basis of laboratory findings [24]. The WFS (wrapper feature selected) algorithm was integrated with 4 diverse classification algorithms named CNN (Convolutional Neural Network), DT, KNN and NB to predict COVID-19. The real dataset was gathered from one hospital of Brazil. The obtained results revealed that the suggested approach performed more efficiently in comparison with other techniques and yielded the accuracy around 76%.

MiodragZivkovic, et.al (2020) recommended a hybridized approach in which ML (machine learning) technique named ANFIS was integrated with enhanced BAS metaheuristics [25]. The enhanced BAS was adopted for determining the parameters of the ANFIS. First of all, the limitations of the traditional approach were tackled by an enhanced BAS algorithm. Thereafter, the data collected from WHO was utilized for quantifying the recommended approach in order to predict COVID-19. The simulation results exhibited that the recommended approach provided superior performance against traditional techniques and proved as a valuable tool to achieve the time-series prediction.

K. Shyam Sunder Reddy, et.al (2020) investigated a NN (neural network)-based technique to predict a number of cases in India infected with Coronavirus Disease 2019 [26]. The LSTM based on RNN was implemented on India dataset to predict the corona disease. A comparative analysis was conducted on two parameter reduced variants of long short term

memory which were generated when the hidden unit signals, bias and input signal were removed. The investigated technique was quantified with regard to MSE (mean square error).The MSE obtained from the investigated technique was calculated 0.59 while predicting the number of cases of corona daily and suggested for taking proper precautions for alleviating the transmission of disease.

Research Gaps

The gaps in this work are discussed below:

1. The dataset for predicting COVID-19 outbreak is complex in nature. Therefore, relationships between different features can't be established easily. Identifying key features for prediction of COVID-19 is troublesome when there is no proper relationship between features.

2. The research work in the past did not apply any outlier removal method for improving the prediction efficiency. Outlier removal will clean up and classify the dataset for advanced processing.

3. The pre-processing stage should present standard methods such as under-sampling or oversampling for the dataset cleaning. Proper cleaning of the dataset improves the efficiency of the prediction analysis process.

Chapter 3

Problem Formulation and Objectives

3.1. Problem Formulation

The essence of this research work is the prediction of covid-19 using data mining. There are three fundamental steps of the COVID-19 prediction technology. The pre-processing phase involves the segregation of missing, superfluous values from the dataset. The next step develops the relationship between the attribute and the target set. The final step is concerned with the division of the entire data into two subsets for training and testing. To predict Covid-19 disease, this work implements random Forest, C4.5 and Multilayer Perceptron. The ensemble classifier gets input as output generated by these three classifier models for COVID-19 prediction. There are many performance metrics such as sensitivity, precision, recall and F1 measure to examine the performance of the ensemble classifier framework. It is required to lessen the complexity of the ensemble classifier for COVID-19 prediction.

3.2 Objectives

The key objectives of this research work are discussed as below:

1. To Implement and analyze various Machine learning Covid-19 prediction techniques.
2. To Design Ensemble scheme for Covid-19 Prediction.
3. To Optimize Voting Regression Model for the analysis of Time Series Datasheet.
4. To compare implemented model with accuracy, Precision, Recall, MSE and RMSE etc.

Chapter 4

Research Methodology

The focus of this research work is the prediction of Covid-19 disease. There are various stages of Covid-19 prediction analysis, explained as follow:

1. **Dataset input and pre-processing:** -The first stage accepts data as input from a trustworthy source. The pre-processed operations based on under-sampling and over-sampling are applied on the dataset in order to remove missing and superfluous values.

2. **Feature extraction and outlier removal**: -The next stage of feature extraction and outlier removal develops relationship between attribute and target set. The feature extraction stage involves identifying key features of the dataset. The outliers are removed using a clustering approach called DBSCAN to clean up data.

3. **Feature Reduction and Classification:** -In this step, PCA algorithm is implemented for feature reduction. The PCA algorithm finds the main features for the next step of classification. The classification approach will be implemented in the third phase. This stage applies the voting ensemble classification. The voting ensemble is the unification of three classifier models, namely logistic regression, random forest and SVM.

4. **Performance Analysis:** -This is the final stage which is concerned with the analysis of the new approach over some universal metrics such as accuracy, precision and recall.

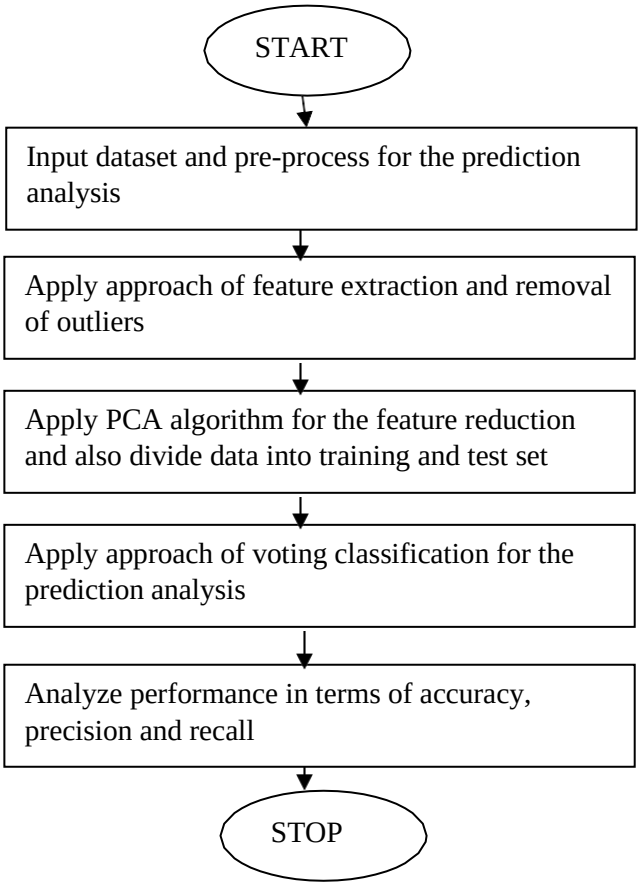
START
Input dataset and pre-process for the prediction analysis
Apply approach of feature extraction and removal of outliers
Apply PCA algorithm for the feature reduction and also divide data into training and test set
Apply approach of voting classification for the prediction analysis
Analyze performance in terms of accuracy, precision and recall
STOP

Figure 4.1 Proposed Methodology

4.1. Dataset Description

The dataset to be used contains the information of positive and negative cases which is reported by the General Directorate of Epidemiology, Secretariat of Health in Mexico. This dataset is generated from the reported results of the RT-PCR test. The dataset contains approx. 600 instances. The dataset is split into training set and testing set in the ratio of 70:30.

.

Chapter 5

Scope of Study and Possible Outcomes

5.1. Scope of Study

This research work has a wide scope in the context of predictive analysis of covid-19. The existent schemes of COVID-19 prediction have many downsides. These downsides are associated with feature extraction, outlier removal and classification. The key purpose of this research work is to address the existing deficits. The elimination of these disadvantages will scale up the accuracy of Covid-19 prediction.

5.2. Expected Outcomes

Following outcomes are likely to be accepted from this research work: -

1. The developed scheme will not only isolate outliers from the dataset but also develop a good relationship between features.

2. The developed scheme will also perform better in the context of accuracy, precision and recall.

References

[1] Yifan Yang, Wenwu Yu, Duxin Chen, “Prediction of COVID-19 spread via LSTM and the deterministic SEIR model”, 2020, 39th Chinese Control Conference (CCC)

[2] Zainab Abbas AbdulhusseinAlwaeli, Abdullahi Abdu Ibrahim, “Predicting Covid-19 Trajectory Using Machine Learning”, 2020, 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)

[3] Suraj Bodapati, Harika Bandarupally, M Trupthi, “COVID-19 Time Series Forecasting of Daily Cases, Deaths Caused and Recovered Cases using Long Short Term Memory Networks”, 2020, IEEE 5th International Conference on Computing Communication and Automation (ICCCA)

[4] Hanqing Chao, Xi Fang, Pingkun Yan, “Integrative analysis for COVID-19 patient outcome prediction”, 2020, Medical Image Analysis

[5] Danish Rafiq, Suhail Ahmad Suhail, Mohammad Abid Bazaz, “Evaluation and prediction of COVID-19 in India: A case study of worst hit states”, 2020, Chaos, Solitons & Fractals

[6] Farah Shahid, AneelaZameer, Muhammad Muneeb, “Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM”, 2020, Chaos, Solitons & Fractals

[7] Anwar Jarndal, Saddam Husain, Omar Zaatar, Talal Al Gumaei, Amar Hamadeh, “GPR and ANN based Prediction Models for COVID-19 Death Cases”, 2020, International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)

[8] Pratima Kumari, Durga Toshniwal, “Real-time estimation of COVID-19 cases using machine learning and mathematical models - The case of India”, 2020, IEEE 15th International Conference on Industrial and Information Systems (ICIIS)

[9] Yasin Khan, Pritam Khan, Sudhir Kumar, Jawar Singh, Rajesh M. Hegde, “Detection and Spread Prediction of COVID-19 from Chest X-ray Images using Convolutional Neural Network-Gaussian Mixture Model”, 2020, IEEE 17th India Council International Conference (INDICON)

[10] Choujun Zhan, Yufan Zheng, Haijun Zhang, Quansi Wen, “Random-Forest-Bagging Broad Learning System with Applications for COVID-19 Pandemic”, 2021, IEEE Internet of Things Journal

[11] Yifan Yang, Wenwu Yu, Duxin Chen, “Prediction of COVID-19 spread via LSTM and the deterministic SEIR model”, 2020, 39th Chinese Control Conference (CCC)

[12] Leonardo Sestrem de Oliveira, Sarah Beatriz Gruetzmacher, João Paulo Teixeira, “COVID-19 Time Series Prediction”, 2021, Procedia Computer Science

[13] Durga Prasad Kavadi, Rizwan Patan, Amir H. Gandomi, “Partial derivative Nonlinear Global Pandemic Machine Learning prediction of COVID 19”, 2020, Chaos, Solitons & Fractals

[14] Ankan Ghosh Dastider, Farhan Sadik, Shaikh Anowarul Fattah, “An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound”, 2021, Computers in Biology and Medicine

[15] Abhishek Dixit, Ashish Mani, Rohit Bansal, “CoV2-Detect-Net: Design of COVID-19 prediction model based on hybrid DE-PSO with SVM using Chest X-ray images”, 2021, Information Sciences

[16] Elena Casiraghi, Dario Malchiodi, Gabriella Trucco, Marco Frasca, Luca Cappelletti, Tommaso Fontana, Alessandro Andrea Esposito, Emanuele Avola, Alessandro Jachetti, Justin Reese, Alessandro Rizzi, Peter N. Robinson, Giorgio Valentini, “Explainable Machine Learning for Early Assessment of COVID-19 Risk Prediction in Emergency Departments”, 2020, IEEE Access

[17] Anwar Jarndal, Saddam Husain, Omar Zaatar, Talal Al Gumaei, Amar Hamadeh, “GPR and ANN based Prediction Models for COVID-19 Death Cases”, 2020, International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)

[18] Huan Zhao, Yichi Li, Shenglan Chu, Suling Zhao, Chenglin Liu, “A COVID-19 Prediction Optimization Algorithm Based on Real-time Neural Network Training—Taking Italy as an Example”, 2021, IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)

[19] Nanning Zheng, Shaoyi Du, Jianji Wang, He Zhang, Wenting Cui, Zijian Kang, Tao Yang, Bin Lou, Yuting Chi, Hong Long, Mei Ma, Qi Yuan, Shupei Zhang, Dong Zhang, Feng Ye, Jingmin Xin, "Predicting COVID-19 in China Using Hybrid AI Model", 2020, IEEE Transactions on Cybernetics

[20] Safa Bahri, MoetezKdayem, NesrineZoghlami, "Deep Learning for COVID-19 prediction", 2020 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)

[21] Teddy Mantoro, RahmadyaTriasHandayanto, Media Anugerah Ayu, Jelita Asian, "Prediction of COVID-19 Spreading Using Support Vector Regression and Susceptible Infectious Recovered Model", 2020, 6th International Conference on Computing Engineering and Design (ICCED)

[22] B. Prabha, Sandeep Kaur, HarikumarPallathadka, "Intelligent predictions of Covid disease based on lung CT images using machine learning strategy", 2021, Materials Today: Proceedings

[23] Wencheng Zou, "The COVID-19 Pandemic Prediction in the US Based on Machine Learning", 2020, International Conference on Public Health and Data Science (ICPHDS)

[24] Hamza Turabieh, Wahiba Ben AbdessalemKaraa, "Predicting the Existence of COVID-19 using Machine Learning Based on Laboratory Findings", 2021, International Conference of Women in Data Science at Taif University (WiDSTaif)

[25] MiodragZivkovic, Nebojsa Bacanin, Fadi Al-Turjman, "COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach", 2020, Sustainable Cities and Society

[26] K. Shyam Sunder Reddy, Y. C. A. Padmanabha Reddy, Ch. Mallikarjuna Rao, "Recurrent neural network based prediction of number of COVID-19 cases in India", 2020, Materials Today: Proceedings

Curiginal

Document Information

Analyzed document	deepti phd synopsis.pdf (D112847633)
Submitted	9/18/2021 4:09:00 AM
Submitted by	
Submitter email	dean-rrc@mits.ac.in
Similarity	8%
Analysis address	dean-rrc.mits@analysis.urkund.com

Sources included in the report

W	URL: https://www.researchgate.net/publication/342410608_COVID-19_Patient_Health_Prediction_Using_Boosted_Random_Forest_Algorithm Fetched: 5/6/2021 2:37:15 PM	1
W	URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8138040/ Fetched: 9/18/2021 4:09:00 AM	2
W	URL: https://www.semanticscholar.org/author/Yichi-Li/2110464246 Fetched: 9/18/2021 4:09:00 AM	1
W	URL: https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019- ncov/resource/en/covidwho-1142827 Fetched: 9/18/2021 4:09:00 AM	2
W	URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8359727/ Fetched: 8/23/2021 12:23:19 PM	2
W	URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8256231/ Fetched: 7/29/2021 5:29:33 AM	1
W	URL: https://ojs.unimal.ac.id/mrbj/article/view/3788 Fetched: 9/18/2021 4:09:00 AM	2
W	URL: https://www.cureus.com/articles/30110-closure-of-universities-due-to-coronavirus-disease-2019-covid-19-impact-on-education-and-mental-health-of-students-and-academic-staff/metrics Fetched: 9/18/2021 4:09:00 AM	1
W	URL: https://ejournal.ikado.ac.id/index.php/teknika/article/download/328/141/ Fetched: 6/21/2021 6:05:24 PM	1
W	URL: https://www.hindawi.com/journals/jcse/2021/8785636/ Fetched: 8/5/2021 4:07:52 PM	1
W	URL: https://www.researchgate.net/publication/346078932_Applications_of_Artificial_Intelligence_in_Battling_Against_Covid-19_A_Literature_Review Fetched: 5/6/2021 2:37:11 PM	2
W	URL: https://www.dovepress.com/artificial-intelligence-in-predicting-clinical-outcome-in-covid-19-pat-peer-reviewed-fulltext-article-RMI	1

Fetched: 9/18/2021 4:09:00 AM

URL: https://research.ebscomedical.com/eds?query=Novel%C2%A0coronavirus&searchfield=SU
Fetched: 9/18/2021 4:09:00 AM
1

URL: https://jesaun.journals.ekb.eg/article_174737.html
Fetched: 9/18/2021 4:09:00 AM
1